Skin Lesion Diagnosis

Instructor:

Prof. Puneet Goyal

Team members:

Pradeep Kumar 2019CSM1008

Sam Zabdiel Samuel 2019CSM1012

Challenge Organizers

Sponsors: Canfield Scientific

Clinical Chair: Harald Kittler, M.D. Medical University of Vienna, Vienna, Austria

Computer Vision Chair: Noel C. F. Codella, Ph.D. IBM Research, New York, USA

Challenge Co-Chairs

- **M. Emre Celebi, Ph.D.** *University of Central Arkansas, Conway, Arkansas*
- **Kristin Dana, Ph.D.** *Rutgers University, New Jersey, USA*
- **David Gutman, Ph.D.** *Emory University, Atlanta, USA*
- *And their team members*

Credits: https://challenge2018.isic-archive.com/

About ISIC

The International Skin Imaging Collaboration (ISIC) is an international effort to improve melanoma diagnosis, sponsored by the International Society for Digital Imaging of the Skin (ISDIS).

The ISIC Archive contains the largest publicly available collection of quality controlled dermoscopic images of skin lesions.

Presently, the ISIC Archive contains over 13,000 dermoscopic images, which were collected from leading clinical centers internationally and acquired from a variety of devices within each center.

Credits: https://challenge2018.isic-archive.com/

The Challenge

Skin cancer is diagnosed visually, beginning with an initial clinical screening and followed potentially by dermoscopic analysis, a biopsy and histopathological examination.

Automated classification of skin lesions using images is a challenging task owing to the fine-grained variability in the appearance of skin lesions.

Different types of skin cancer

- Squamous cell carcinoma
- Melanocytic nevi
- Dermatofibrosarcoma protuberans (DFSP)
- Melanoma
- Benign keratosis-like lesions
- Merkel cell carcinoma
- Sebaceous carcinoma
- Basal cell carcinoma
- Actinic keratoses
- Vascular lesions
- Dermatofibroma
- Cutaneous T-cell lymphoma

About Dataset

- We used the HAM10000 dataset.
- A large collection of multi-source dermatoscopic images of common pigmented skin lesions acquired and stored by different modalities and from different population.
- The dataset contains **10015 images of seven skin cancer type** that includes:

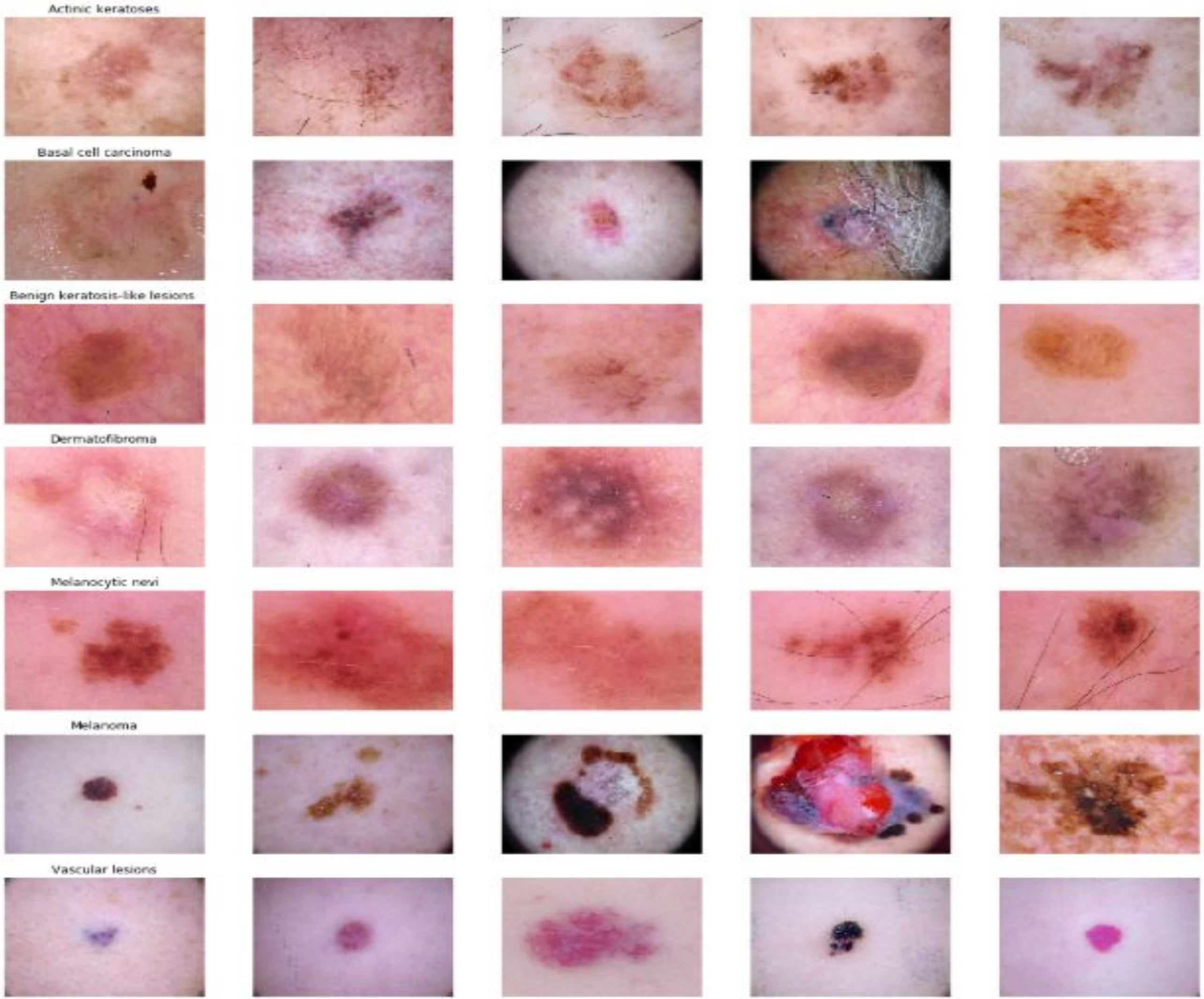
1. Melanocytic nevi (nv)
2. Melanoma (mel)
3. Benign keratosis-like lesion (bkl),
4. Dermatofibroma (df)
5. Basal cell carcinoma (bcc)
6. Actinic keratoses/ Bowen's disease (akiec)
7. Vascular(vasc)

Sample Data

Sequence:

1. akiec
2. bcc
3. Bkl
4. df
5. nv
6. mel
7. vasc

Images size:
600x450 pixels



Challenge details

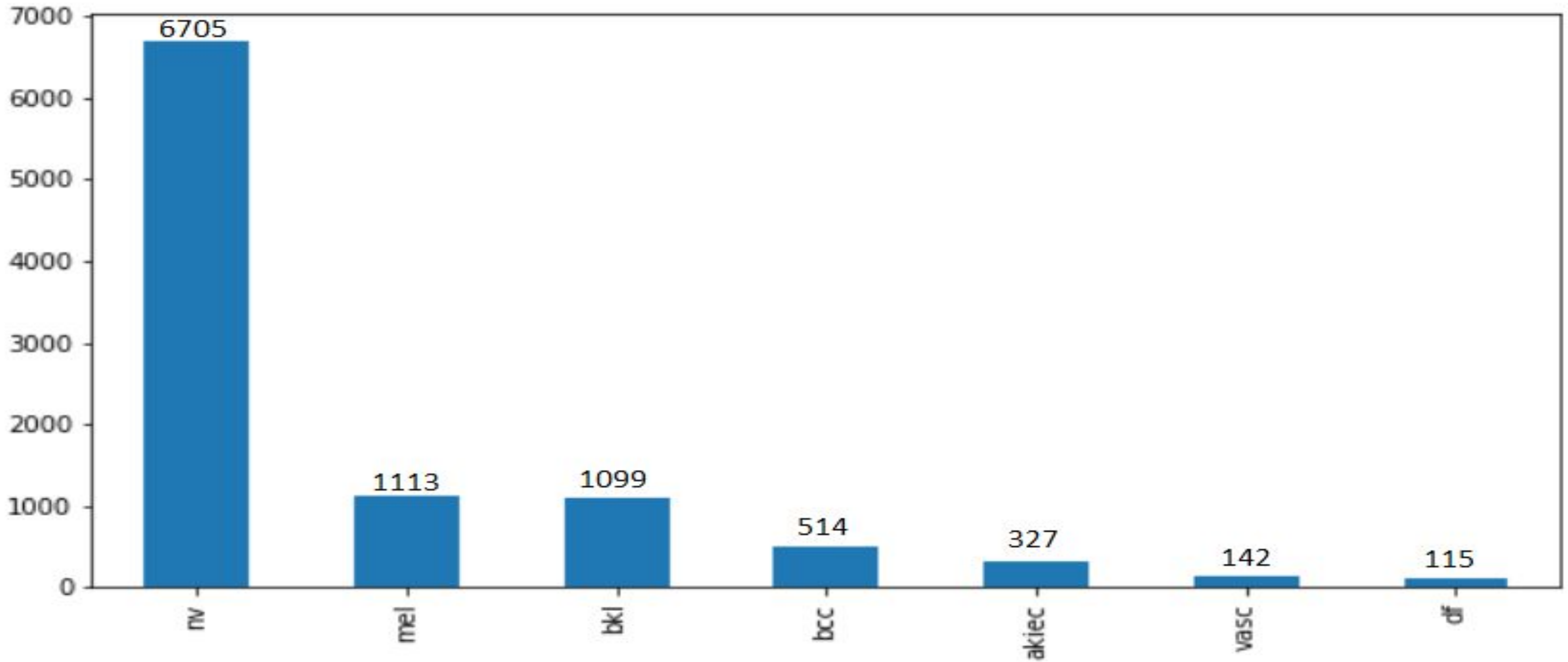
Input: Dermoscopic images to classify into the 7 skin cancer categories.

Output: CSV file giving image file name, Diagnosis confidences are expressed between [0.0, 1.0] for the 7 classes.

Evaluation metric: Predicted responses are scored using a normalized multi-class accuracy metric (balanced across categories). Tied positions will be broken using the area under the receiver operating characteristic curve (AUC) metric. Other metrics like sensitivity, specificity, accuracy, AUC-ROC, mean avg. precision, f1 score, etc were also computed for Scientific completeness.

Data Distribution of each class

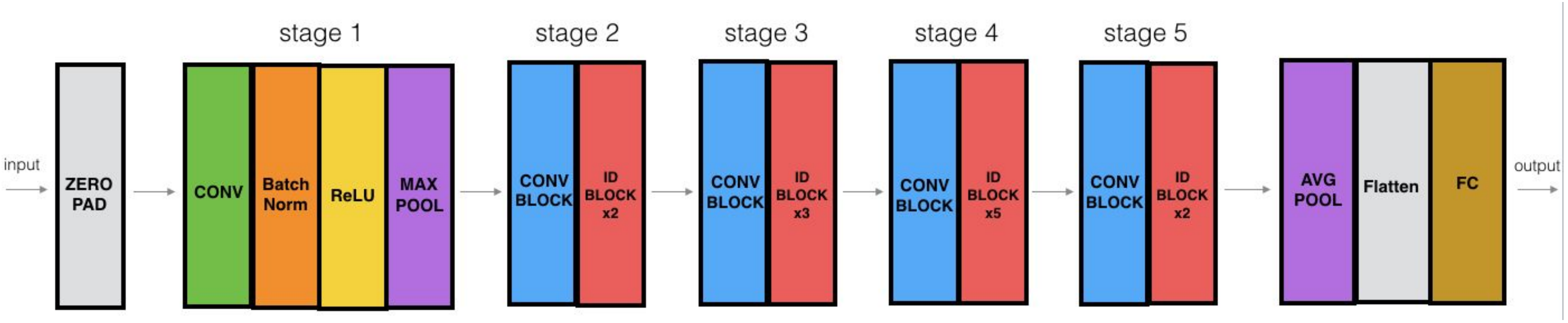
```
<matplotlib.axes._subplots.AxesSubplot at 0x1721e3681c8>
```



Implemented Approach

ResNet-50

- Image size: (100, 75) RGB
- No. of epochs = 50
- Batch size = 10
- No. of layers = 50 + 8 (Conv2d with dropout and dense layers)
- Size of training data = 8012
- Size of validation data = 2003

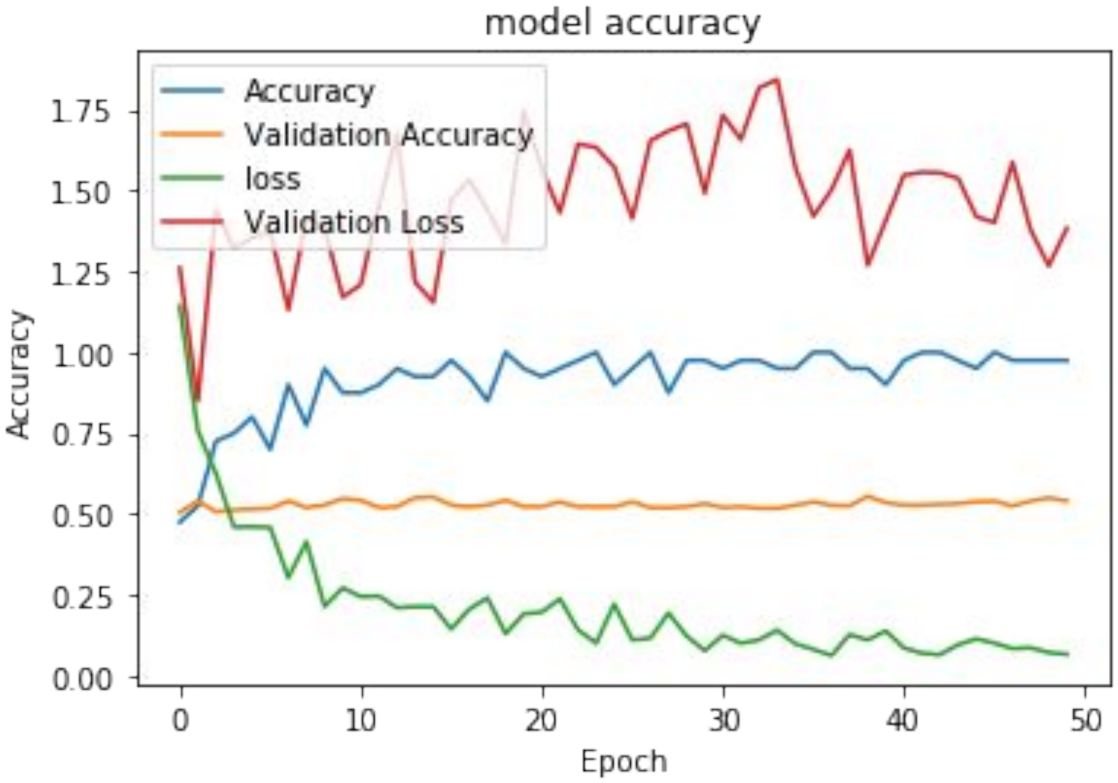


Conv2d added layers architecture

```
Model: "sequential_4"
_________________________________________________________________
Layer (type)                 Output Shape              Param #   
=================================================================
conv2d_25 (Conv2D)           (None, 75, 100, 32)       896       
_________________________________________________________________
conv2d_26 (Conv2D)           (None, 75, 100, 32)       9248      
_________________________________________________________________
max_pooling2d_9 (MaxPooling2 (None, 37, 50, 32)        0         
_________________________________________________________________
dropout_8 (Dropout)          (None, 37, 50, 32)        0         
_________________________________________________________________
conv2d_27 (Conv2D)           (None, 37, 50, 64)        18496     
_________________________________________________________________
conv2d_28 (Conv2D)           (None, 37, 50, 64)        36928     
_________________________________________________________________
max_pooling2d_10 (MaxPooling (None, 18, 25, 64)        0         
_________________________________________________________________
dropout_9 (Dropout)          (None, 18, 25, 64)        0         
_________________________________________________________________
flatten_2 (Flatten)          (None, 28800)             0         
_________________________________________________________________
dense_3 (Dense)              (None, 128)               3686528   
_________________________________________________________________
dropout_10 (Dropout)         (None, 128)               0         
_________________________________________________________________
dense_4 (Dense)              (None, 7)                 903       
=================================================================
Total params: 3,752,999
Trainable params: 3,752,999
Non-trainable params: 0
```

ResNet-50 Results

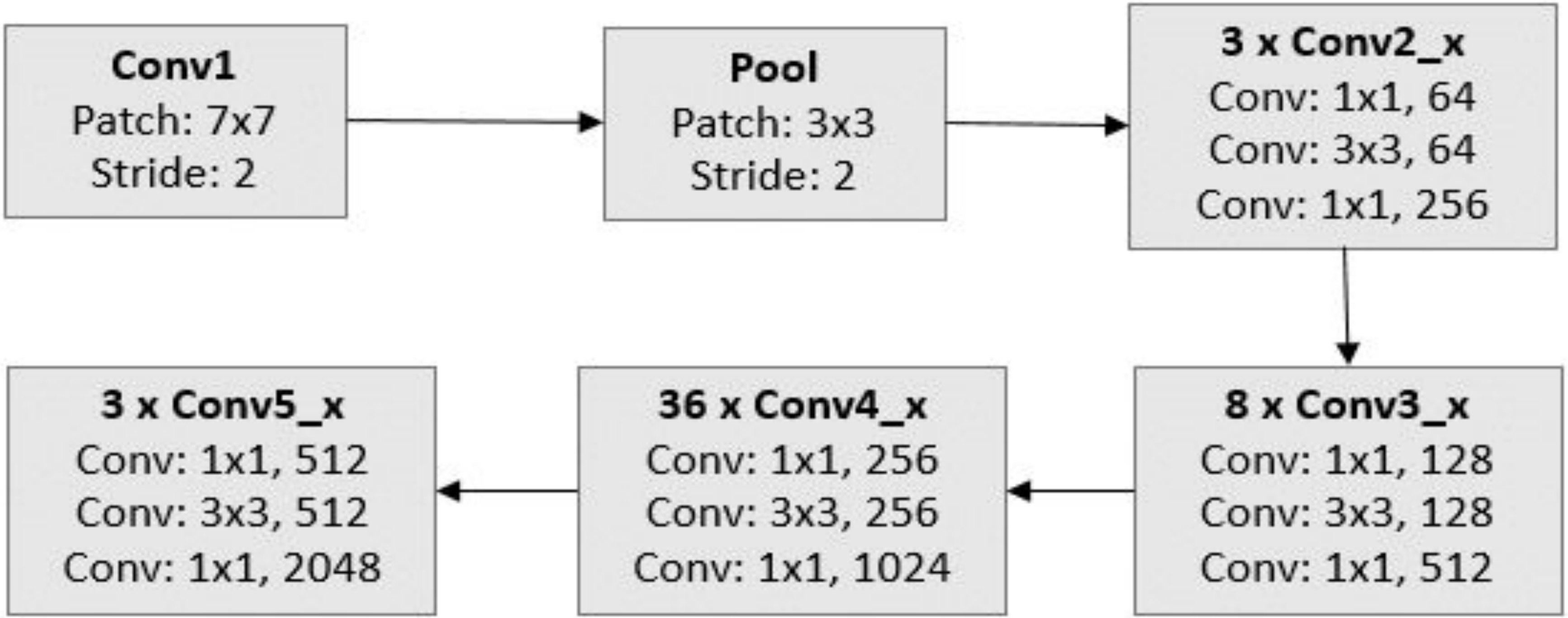
Training Accuracy: 0.975
Validation Accuracy: 0.54
Loss: 0.066
Validation Loss: 1.383



ResNet-152 V2

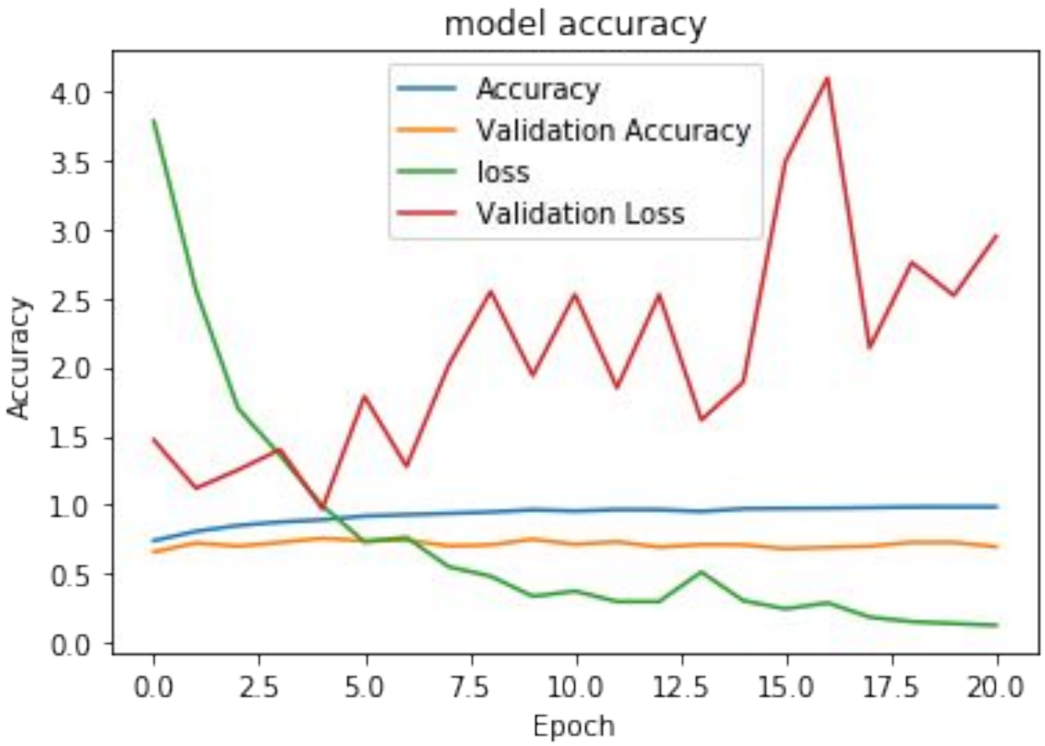
- Image size: (224, 224)RGB
- No. of epochs = 100
- Batch size = 64
- No. of layers = 152 + 2(Global Average Pooling 2D and dense layers)
- Training data size = 7508 images
- Validation data size = 504 images
- Class weights were assigned according to class strength to handle imbalance in data.

Basic architecture of ResNet-152



ResNet-152 V2 Results

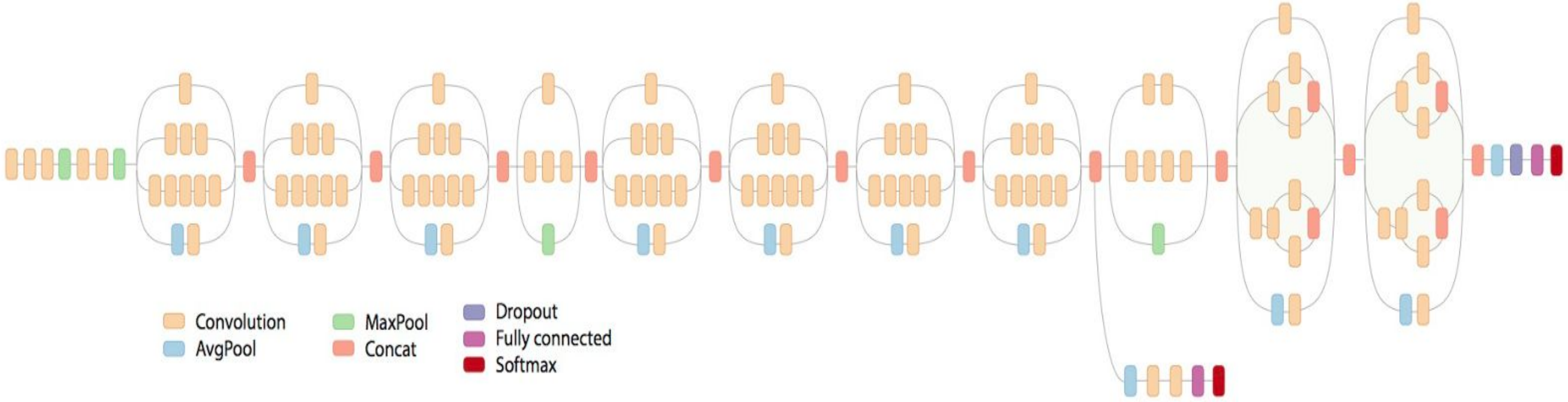
Training Accuracy: 0.9852
Validation Accuracy: 0.6955
Training Loss: 0.126
Validation Loss: 2.9474



Inception V3 model

- Transfer learning is a machine learning method which utilizes a pre-trained neural network. For example, the image recognition model called Inception-v3.
- Image size: (224, 224) RGB
- No. of epochs = 20
- Batch size = 64
- Size of training data = 35538
- Size of validation data = 404
- No. of layers = 50 + 8 (Conv2d with dropout and dense layers)
- Upsampling was done to handle data imbalance. The data from all classes having fewer than 4.5k images were upsampled. All classes had between 4.5k-5k images at least. Augmentations were done to avoid overfitting.

Inception V3 model architecture

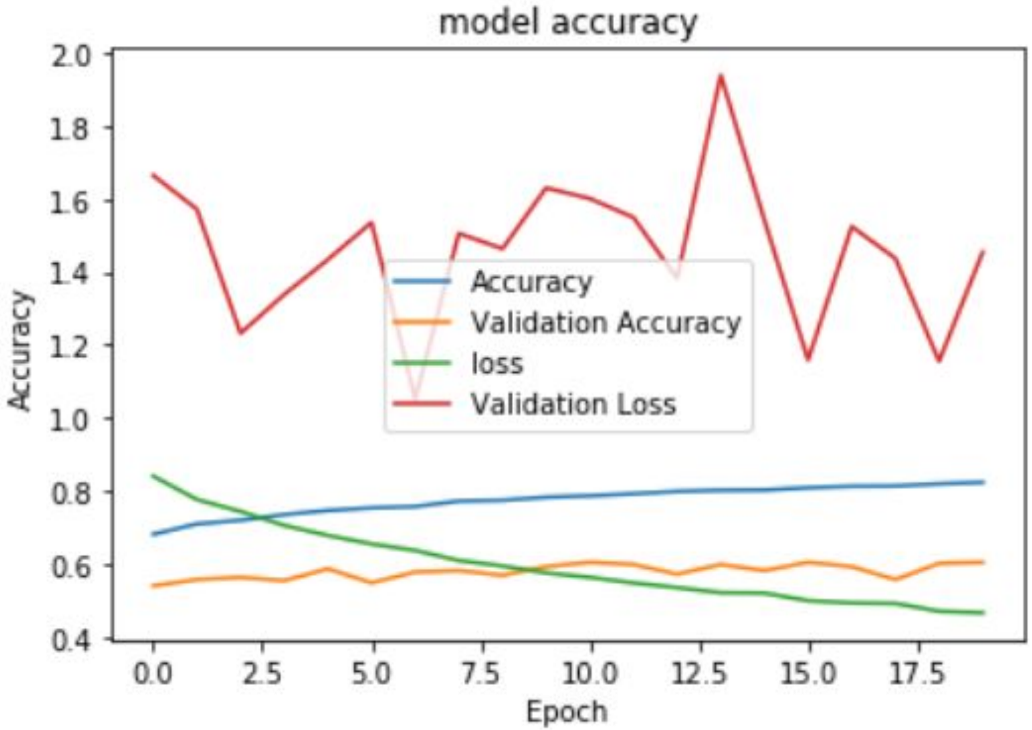


Inception-v3 consists of two parts:

- Feature extraction part with a convolutional neural network.
- Classification part with fully-connected and softmax layers

Inception V3 model results

Training Accuracy: 0.8244
Validation Accuracy: 0.6059
Loss: 0.468
Validation Loss: 1.4541



DenseNet-151 (with weight decay) using Fastai

- Image Size: (128, 128) RGB
- Images normalized, augmented.
- Fastai uses Cyclical Learning Rates policy, which instead of monotonically decreasing learning rate, varies it between reasonable boundary values, leading to faster convergence.
- Weight decay or L2 regularization to avoid overfitting.
- The model gave a balanced accuracy of about ~0.8-0.82, on a set of 5% images withheld from the training dataset for validation.

DenseNet-151 Results

- These results are on validation data.
- Classes = [MEL', 'NV', 'BCC', 'AKIEC', 'BKL', 'DF', 'VASC']

	MEL	NV	BCC	AKIEC	BKL	DF	VASC
MEL	40	15	1	0	5	0	0
NV	5	321	1	1	2	0	1
BCC	0	0	23	0	2	0	0
AKIEC	2	0	0	8	4	0	0
BKL	4	5	1	1	42	0	0
DF	0	1	0	0	0	5	0
VASC	0	2	0	0	0	0	8

Confusion matrix

	precision	recall	f1-score	support
0	0.78	0.66	0.71	61
1	0.93	0.97	0.95	331
2	0.88	0.92	0.90	25
3	0.80	0.57	0.67	14
4	0.76	0.79	0.78	53
5	1.00	0.83	0.91	6
6	0.89	0.80	0.84	10
accuracy			0.89	500
macro avg	0.86	0.79	0.82	500
weighted avg	0.89	0.89	0.89	500

Classification report

Results at a glance

Method	Training Accuracy	Testing/Validation Accuracy
ResNet-50	0.975	0.54
Inception V3	0.824	0.60
ResNet-152 V2	0.98	0.70
DenseNet-151	-	0.89

Conclusion

ResNet-50 overfits with validation accuracy only 0.54. ResNet-152 also seems to overfit the training data with large difference between training and testing accuracy. Fine tuning was also unsuccessful on Inception V3 model.

DesnseNet-151 with weight decay regularizer has outperformed other models. It has an accuracy of ~0.87-0.90 and balanced multi-class accuracy of ~0.8-0.82 on validation data.

References

Our data was extracted from the “ISIC 2018: Skin Lesion Analysis Towards Melanoma Detection” grand challenge datasets [1][2].

[1] Tschandl P., Rosendahl C. & Kittler H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci. Data 5, 180161 doi.10.1038/sdata.2018.161 (2018)

[2] Noel Codella, Veronica Rotemberg, Philipp Tschandl, M. Emre Celebi, Stephen Dusza, David Gutman, Brian Helba, Aadi Kalloo, Konstantinos Liopyris, Michael Marchetti, Harald Kittler, and Allan Halpern: “Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC)”, 2018; arXiv:1902.03368.

[3] L. N. Smith, "Cyclical Learning Rates for Training Neural Networks," 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), Santa Rosa, CA, 2017, pp. 464-472, doi: 10.1109/WACV.2017.58.

References

[4] Skin Lesion Analysis Towards Melanoma Detection via End-to-end Deep Learning of Convolutional Neural Networks by Katherine M. Li and Evelyn C. Li (Leaderboard position: 7)

[5] Skin Lesion Analysis Towards Melanoma Detection Using Deep Neural Network Ensemble by Jiaxin Zhuang, Weipeng Li , Siyamalan Manivannan from University of Dundee and University of Jaffna (Leaderboard position: 15)

[6] Residual Network based Aggregation Model for Skin Lesion Classification by Yongsheng Pan1 Yong Xia1 from Northwestern Polytechnical University, Xi'an, China (Leaderboard position: 17)


Thank you!